5.4 The spatial error model
5.4.1 Regression model with spatially autocorrelated errors

In a **multiple regression model**, the dependent variable Y depends on k regressors X_1(=1) , X_2, …, X_k and a disturbance ε:

(4.1) $$\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

y is a nx1 vector of the geo-referenced endogenous variable Y, **X** a nxk observation matrix of the regressors and **ε** a nx1 disturbance vector. However, different from the standard regression model (see section 4.1), the disturbances ε_i are not assumed to be independently identically distributed (i.i.d.). Instead ε_i is assumed to follow a spatially autocorrelated process.

In principle, the error process can be modelled by a **spatially autoregressive process**,

(5.35) $$\boldsymbol{\varepsilon} = \lambda \mathbf{W} \boldsymbol{\varepsilon} + \mathbf{v} \quad ,$$

or a **spatially moving average process**, where **ε** in (4.1) is replaced by

(5.36) $$\boldsymbol{\varepsilon} = \mathbf{v} + \lambda \mathbf{W} \mathbf{v} \;.$$

W is a nxn spatial weight matrix, **v** a n+1 vector i.i.d. random variables and λ a spatially autoregressive parameter.

The model (4.1) with the error processes (5.35) or (5.36) is called a **regression model with spatially autocorrelated disturbances**. Both spatial error processes are first-order processes. Usually the error dependence is modelled by (5.35). This modelling approach is also adopted here. The model (4.1) with the error process (5.35) is called **regression model with spatially autoregressive disturbances**.

When the **spatial error model** (4.1) and (5.35) is estimated by the **method of maximum likelihood**, the **normal assumption** for the error term $\mathbf{v}$ must be imposed:

(5.37) $\mathbf{v} \sim \mathrm{N}(0, \sigma^2 \mathbf{I})$

As the error process (5.35) can be written in the form

$$\boldsymbol{\varepsilon} = \boldsymbol{\lambda} \mathbf{W} \boldsymbol{\varepsilon} + \mathbf{v}$$

$$(\mathbf{I} - \lambda \mathbf{W}) \cdot \boldsymbol{\varepsilon} = \mathbf{v}$$

(5.38) $\boldsymbol{\varepsilon} = (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{v}$

the **spatial error model** (4.1) with (5.35) is given in dependence of the i.i.d. random term $\mathbf{v}$ (=data generated process) by

(5.39) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda \mathbf{W})^{-1} \mathbf{v}$

From econometric analysis of time-series it is well-known that autocorrelated errors will not affect the property of unbiasedness of the OLS estimator for **β** in the standard regression model. The same holds in case of **spatial error auto-correlation**, where **OLS** still provides **unbiased** estimates of the regression coefficients.

However, the regression coefficients are **inefficiently** estimated as their **standard errors** are **biased**. OLS only provides efficient parameter estimates if the covariance matrix of the errors ε_i is a scalar matrix ($\sigma^2 \cdot$ **I** in the standard regression model, see section 4.1). For the error term (5.29), however, the covariance matrix reads

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') &= \mathrm{E}\{[(\mathbf{I}-\lambda\mathbf{W})^{-1}\mathbf{v}][(\mathbf{I}-\lambda\mathbf{W})^{-1}\mathbf{v}]'\} \\ &= \mathrm{E}\{(\mathbf{I}-\lambda\mathbf{W})^{-1}\mathbf{v}\mathbf{v}'[(\mathbf{I}-\lambda\mathbf{W})^{-1}]'\} \\ &= (\mathbf{I}-\lambda\mathbf{W})^{-1}\sigma^2\mathbf{I}[(\mathbf{I}-\lambda\mathbf{W})^{-1}]'\end{aligned}$$

or

(5.40) $$\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2(\mathbf{I}-\lambda\mathbf{W})^{-1}[(\mathbf{I}-\lambda\mathbf{W}')^{-1}]$$

In econometric models with non-scalar covariance matrices, **generalised least-squares (GLS)** is usually recommended for efficient parameter estimation.

The inefficiency of OLS estimates of the regression coefficients would invalidate statistical inference in the **spatial error model** (4.1)/(5.35) or (5.39). Particularly, significance tests for OLS estimated regression coefficients are no more valid. The invalidity of significance tests arises from biased estimation of the variances and standard errors of the OLS estimates for $\boldsymbol{\beta}$ and λ.

As already noted, application of the **generalised least-squares (GLS) method** will accomplish efficient parameter estimates. Alternatively, the **method of maximum likelihood (ML)** can be applied. It leads to estimates of the regression coefficients and the error variance that maximise the joint probability density of the given sample. In our exposition, we will adopt the maximum likelihood approach of parameter estimation.

5.4.2 Maximum likelihood estimation in the spatial error model

The spatial error model discussed in section 5.4.1 is given by the equations (4.1) and (5.35),

(4.1) $$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$$

(5.35) $$\boldsymbol{\varepsilon} = \boldsymbol{\lambda}\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{v}$$

or equation (5.43)

(5.39) $$\mathbf{y} = \mathbf{X\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\mathbf{v}$$

The errors v_i are assumed to be independently normally distributed:

(5.37) $$\mathbf{v} \sim \mathrm{N}(0, \sigma^2\mathbf{I})$$

By solving (5.39) for $\mathbf{v}$,

(5.41) $$\mathbf{v} = (\mathbf{I} - \lambda\mathbf{W})(\mathbf{y} - \mathbf{X\beta})$$

we obtain the Jacobian

(5.45) $$\mathrm{J} = \left|\frac{\partial\mathbf{v}}{\partial\mathbf{y}}\right| = \left|\mathbf{I} - \lambda\mathbf{W}\right|$$

Thus, the **log-likelihood function** for **y** of the **spatial error model** is obtained by adding the term ln |**I**-λ**W**| to the log likelihood function of the standard regression model (B5.11) and replacing **u** by $\boldsymbol{\nu}$ according to (5.41):

(5.43a) $$\ln L\left(\boldsymbol{\beta},\lambda,\sigma^2 \mid \mathbf{y},\mathbf{X}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I}-\lambda\mathbf{W}\right| - \frac{1}{2\sigma^2}\boldsymbol{\nu}'\boldsymbol{\nu}$$

Replacing the random vector $\boldsymbol{\nu}$ by (5.41) the log-likelihood function reads

(5.43b) $$\ln L\left(\boldsymbol{\beta},\lambda,\sigma^2 \mid \mathbf{y},\mathbf{X}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I}-\lambda\mathbf{W}\right|$$
$$- \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}-\lambda\mathbf{W})'(\mathbf{I}-\lambda\mathbf{W})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

Maximising the log likelihood function, ln L ,is equal to minimising the sum of transformed squared errors, $\boldsymbol{\nu}'\boldsymbol{\nu}$, corrected by the log of the Jacobian, ln $|\mathbf{I}-\lambda\mathbf{W}|$. On account of this correction the ML estimates will differ from the OLS estimates. They coincide for $\lambda \to 0$ where the spatial error model approaches the standard regression model.

[Asymptotic properties like consistency, asymptotic efficiency and asymptotic normality of the ML estimates will only hold if regularity conditions are satisfied. Particularly the Jacobian J is required to be positive: $J = |\mathbf{I}-\lambda\mathbf{W}| > 0$]

An ML estimator for $\boldsymbol{\beta}$ is obtained by setting the first partial derivative of the loglikelihood function (5.43b) with respect to $\boldsymbol{\beta}$ equal to zero. This estimator is equal with the **generalised least-squares estimator (GLS estimator)**, $\hat{\boldsymbol{\beta}}_{\mathrm{ML}} = \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$, as it can be thought as a least-squares estimator resulting from a regression of **y*** on **X***

(5.44) $\mathbf{y}^* = (\mathbf{I} - \lambda\mathbf{W})\cdot\mathbf{y}$ and (5.45) $\mathbf{X}^* = (\mathbf{I} - \lambda\mathbf{W})\cdot\mathbf{X}$.

when the autoregressive parameter λ is known:

(5.46a) $$\hat{\boldsymbol{\beta}}_{\mathrm{ML}} = \hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$$

or, more detailed,

(5.46b) $$\hat{\boldsymbol{\beta}}_{\mathrm{ML}} = \hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = [\mathbf{X}'(\mathbf{I}-\lambda\mathbf{W})'(\mathbf{I}-\lambda\mathbf{W})\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I}-\lambda\mathbf{W})'(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}$$

[As (**I** - **W**) is the spatial difference operator, the transformation matrix (**I** - λ**W**) can be interpreted as a generalised spatial difference operator. For $\lambda \to 1$ the transformations (5.44) and (5.45) consist on forming spatial differences of the involved variables.]

A first-order condition resulting from the partial derivative of ln L with respect to σ^2 gives the ML estimator for the error variance

(5.47a) $\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}(\mathbf{e}-\lambda\mathbf{W}\mathbf{e})'(\mathbf{e}-\lambda\mathbf{W}\mathbf{e})$ or (5.47b) $\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}\mathbf{e}'(\mathbf{I}-\lambda\mathbf{W})'(\mathbf{I}-\lambda\mathbf{W})\mathbf{e}$

with the residual vector

(5.48) $$\mathbf{e} = \mathbf{y} - \mathbf{X}(\hat{\boldsymbol{\beta}}_{\mathrm{ML}} = \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}) \ .$$

Note, however, that of both ML estimators, $\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{GLS}$ and $\hat{\sigma}^2_{ML}$, are functions of the autoregressive parameter λ. A ML estimator of λ can be found by maxi-mising the **concentrated log-likelihood function**

(5.49) $$\ln L_c(\lambda) = C - \frac{n}{2} \cdot \ln[\frac{1}{n}\mathbf{e}'(\mathbf{I} - \lambda\,\mathbf{W})'(\mathbf{I} - \lambda\,\mathbf{W})\mathbf{e}] + \ln\left|\mathbf{I} - \lambda\,\mathbf{W}\right|$$

that is obtained by inserting the ML estimates (5.46b) and (5.47a/b) for the para-meters $\boldsymbol{\beta}$ and σ^2. C is a constant.

ln L_c is a nonlinear function of λ for which no analytical solution exists. Thus, a ML estimator for λ has to be found by numerical optimisation. However, the residual vector **e** in ln L_c depends indirectly also on λ, as the calculation of $\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{GLS}$ requires a value of λ. Therefore an iteration procedure must be con-ceived where both parameters λ and $\boldsymbol{\beta}$ are updated conditional on one another.

Flowchart of ML estimation in the spatial error model

Step 1: OLS estimation of **y** on **X**
OLS estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

Step 2: Compute OLS residual vector: $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$

Step 3: Search for the autoregressive parameter λ that maximises ln L_c given $\hat{\boldsymbol{\beta}}_{GLS}$ and **e**

Step 4: GLS estimation of **y*** = (**I** - λ**W**)·**y** on **X*** = (**I** - λ**W**)·**X** given λ
GLS estimator (5.46b): $\hat{\boldsymbol{\beta}}_{GLS} = [\mathbf{X}'(\mathbf{I}-\lambda\mathbf{W})'(\mathbf{I}-\lambda\mathbf{W})\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I}-\lambda\mathbf{W})'(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}$

Step 5: Compute GLS residual vector: $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}$

Step 6: Convergence criterion met? — no → Step 3
|
yes → Step 7

Step 7: Compute ML estimator for the error variance given **e** (with $\hat{\boldsymbol{\beta}}_{GLS}$) and λ

ML estimator (5.47b): $\hat{\sigma}^2_{ML} = \frac{1}{n}\mathbf{e}'(\mathbf{I}-\lambda\,\mathbf{W})'(\mathbf{I}-\lambda\,\mathbf{W})\mathbf{e}$

Example:

The data on output growth (X) and productivity growth (Y) of the 5-region example are used to illustrate ML estimation in the spatial error model.

Region	1	2	3	4	5
Output growth (X)	0.6	1.0	1.6	2.6	2.2
Productivity growth (Y)	0.4	0.6	0.9	1.1	1.2

The spatial error model relates the regional productivity growth to own region‘s output growth and spatially autocorrelated disturbances:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i + \lambda \cdot \sum_{j=1}^{n} w_{ij} \cdot \varepsilon_j \qquad (5.50)$$

The spatially lagged error term may capture measurement errors due to inadequate delineation of regions as well as spatial effects generated by unobserved attributes.

ML estimation in the spatial error model

To illustrate ML estimation in the spatial error model, we compute the one-step ML estimates, i.e. we give the ML estimates resulting after one iteration.

Step 1: OLS estimation of **y** on **X** (see section 4.1)

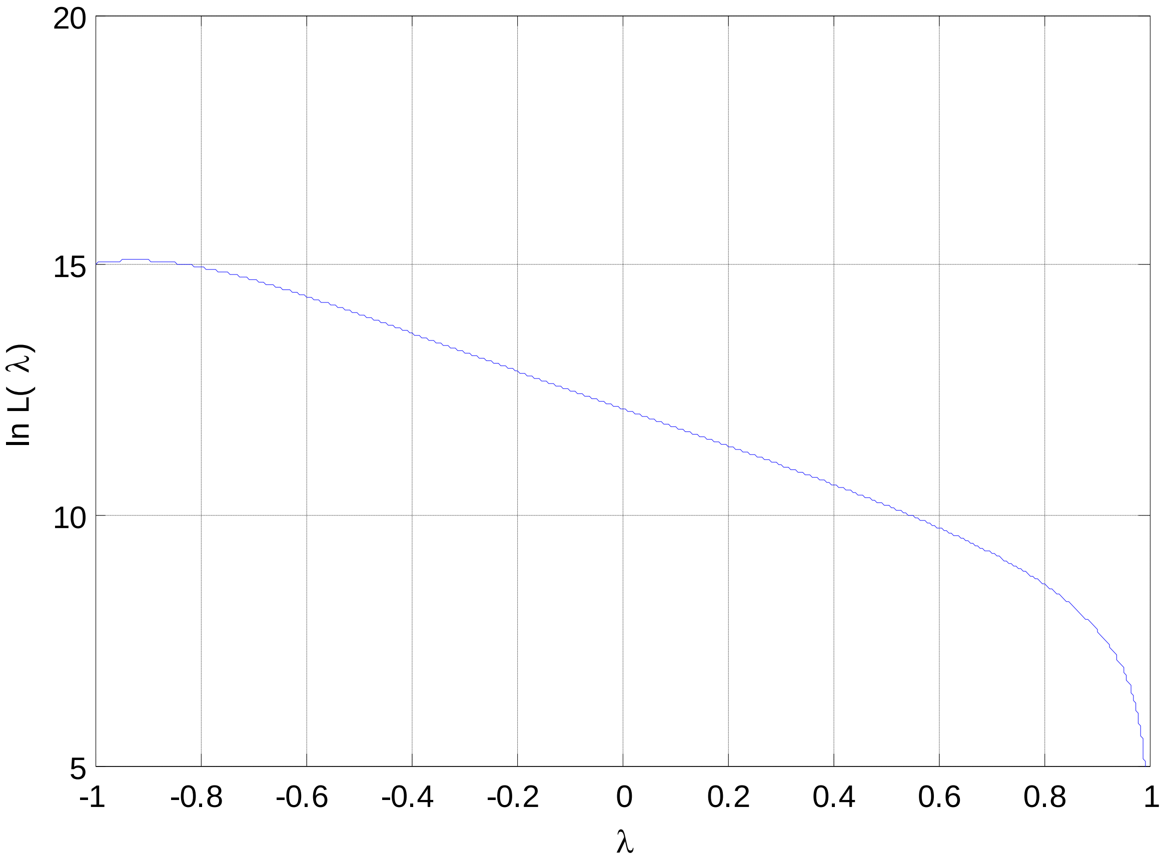
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1.1412 & -0.5882 \\ -0.5882 & 0.3676 \end{bmatrix} \begin{bmatrix} 4.20 \\ 7.78 \end{bmatrix} = \begin{bmatrix} 0.2165 \\ 0.3897 \end{bmatrix}$$

Step 2: Compute OLS residual vector (see section 4.1)

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} - \begin{bmatrix} 0.4503 \\ 0.6062 \\ 0.8400 \\ 1.2297 \\ 1.0738 \end{bmatrix} = \begin{bmatrix} -0.0503 \\ -0.0062 \\ 0.0600 \\ -0.1297 \\ 0.1262 \end{bmatrix}$$

Step 3: Maximisation of the concentrated likelihood function L_C given $\hat{\boldsymbol{\beta}}_{GLS}$ and **e** to obtain an ML estimator $\hat{\lambda}_{ML}$ for the autoregressive parameter λ

Figure: Log L_c function of the extended spatial lag model



ML estimator for λ:

$$\hat{\lambda}_{ML} = -0.9218$$

Log Lc function at $\hat{\lambda}_{ML} = -0.9218$:

$\ln L_c^* = 15.0921$

(neglecting the constant C)

Step 4: GLS estimation of **y*** = (**I** - λ**W**)·**y** on **X*** = (**I** - λ**W**)·**X** using the estimate for λ ($\hat{\lambda}_{ML} = -0.9218$) obtained in step 3

$$\mathbf{X}^* = (\mathbf{I} - \lambda \cdot \mathbf{W}) \cdot \mathbf{X} = \left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + 0.9218 \cdot \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \right) \cdot \begin{bmatrix} 1 & 0.6 \\ 1 & 1.0 \\ 1 & 1.6 \\ 1 & 2.6 \\ 1 & 2.2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -0.4609 & -0.4609 & 0 & 0 \\ -0.3073 & 1 & -0.3073 & -0.3073 & 0 \\ -0.3073 & -0.3073 & 1 & -0.3073 & 0 \\ 0 & -0.3073 & -0.3073 & 1 & -0.3073 \\ 0 & 0 & 0 & -0.9218 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0.6 \\ 1 & 1.0 \\ 1 & 1.6 \\ 1 & 2.6 \\ 1 & 2.2 \end{bmatrix} = \begin{bmatrix} 0.0782 & -0.5983 \\ 0.0782 & -0.4749 \\ 0.0782 & 0.3095 \\ 0.0782 & 1.1251 \\ 0.0782 & -0.1967 \end{bmatrix}$$

$$\mathbf{y}^* = (\mathbf{I} - \lambda \cdot \mathbf{W}) \cdot \mathbf{y} = \left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + 0.9218 \cdot \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \right) \cdot \begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -0.4609 & -0.4609 & 0 & 0 \\ -0.3073 & 1 & -0.3073 & -0.3073 & 0 \\ -0.3073 & -0.3073 & 1 & -0.3073 & 0 \\ 0 & -0.3073 & -0.3073 & 1 & -0.3073 \\ 0 & 0 & 0 & -0.9218 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} = \begin{bmatrix} -0.2913 \\ -0.1374 \\ 0.2547 \\ 0.2704 \\ 0.1860 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\mathbf{X}^{*\prime}\mathbf{y}^{*} = \begin{bmatrix} 32.7949 & -0.2129 \\ -0.2129 & 0.5054 \end{bmatrix} \cdot \begin{bmatrix} 0.0782 & -0.5983 \\ 0.0782 & -0.4749 \\ 0.0782 & 0.3095 \\ 0.0782 & 1.1251 \\ 0.0782 & -0.1967 \end{bmatrix} \cdot \begin{bmatrix} -0.2913 \\ -0.1374 \\ 0.2547 \\ 0.2704 \\ 0.1860 \end{bmatrix}$$

$$= \begin{bmatrix} 32.7949 & -0.2129 \\ -0.2129 & 0.5054 \end{bmatrix} \cdot \begin{bmatrix} 0.0221 \\ 0.5861 \end{bmatrix} = \begin{bmatrix} 0.5993 \\ 0.2915 \end{bmatrix}$$

Step 5: Compute GLS residual vector:

$$\mathbf{e} = \mathbf{y} - \mathbf{X} \cdot \hat{\boldsymbol{\beta}}_{GLS} = \begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} - \begin{bmatrix} 1 & 0.6 \\ 1 & 1.0 \\ 1 & 1.6 \\ 1 & 2.6 \\ 1 & 2.2 \end{bmatrix} \cdot \begin{bmatrix} 0.5993 \\ 0.2993 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} - \begin{bmatrix} 0.7742 \\ 0.8908 \\ 1.0658 \\ 1.3573 \\ 1.2407 \end{bmatrix} = \begin{bmatrix} -0.3742 \\ -0.2908 \\ -0.1658 \\ -0.2573 \\ -0.0407 \end{bmatrix}$$

Step 6: Convergence criterion met?

Only one iteration is carried out for illustrative purposes.

Step 7: Compute ML estimator for the error variance given **e** (with $\hat{\boldsymbol{\beta}}_{GLS}$) and λ

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\mathbf{e}'(\mathbf{I}-\lambda\,\mathbf{W})'(\mathbf{I}-\lambda\,\mathbf{W})\mathbf{e} = \frac{1}{5}\cdot\begin{bmatrix}-0.3742 & -0.2908 & -0.1658 & -0.2573 & -0.0407\end{bmatrix}\cdot$$

$$\begin{bmatrix} 1 & -0.3073 & -0.3073 & 0 & 0 \\ -0.4609 & 1 & -0.3073 & -0.3073 & 0 \\ -0.4609 & -0.3073 & 1 & -0.3073 & 8 \\ 0 & -0.3073 & -0.3073 & 1 & -0.9218 \\ 0 & 0 & 0 & -0.3073 & 1 \end{bmatrix}\cdot$$

$$= \begin{bmatrix} 1 & -0.4609 & -0.4609 & 0 & 0 \\ -0.3073 & 1 & -0.3073 & -0.3073 & 0 \\ -0.3073 & -0.3073 & 1 & -0.3073 & 0 \\ 0 & -0.3073 & -0.3073 & 1 & -0.3073 \\ 0 & 0 & 0 & -0.9218 & 1 \end{bmatrix}\cdot\begin{bmatrix}-0.3742 \\ -0.2908 \\ -0.1658 \\ -0.2573 \\ -0.0407\end{bmatrix} = 0.0185$$

- ML estimated spatial error model of productivity growth (using the one-step estimators):

$$y_i = 0.5953 + 0.2915\cdot x_i + \varepsilon_i - 0.9218\cdot\sum_{j=1}^{n} w_{ij}\cdot\varepsilon_j$$